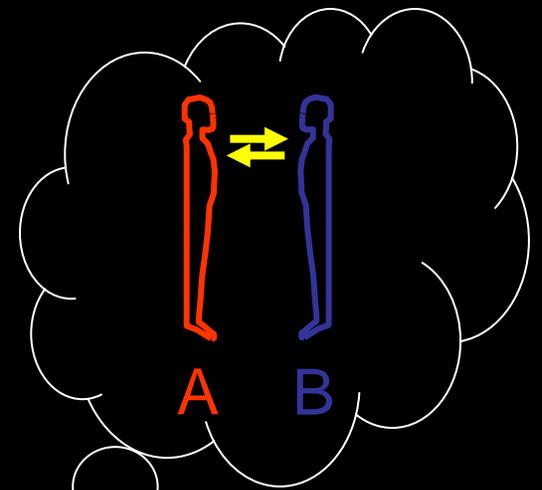
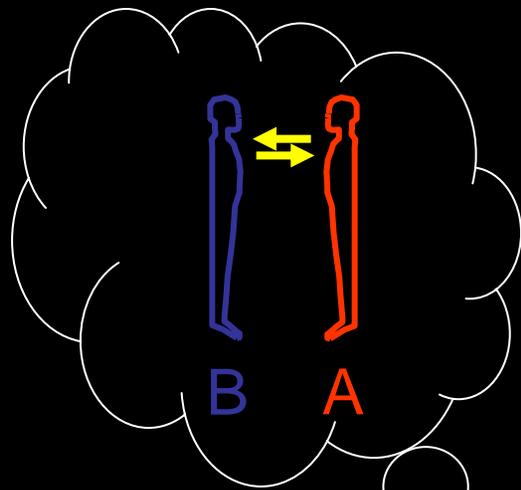


Reward Processing and Social Exchange

Read Montague
Baylor College of Medicine
Houston, TX
www.hnl.bcm.tmc.edu

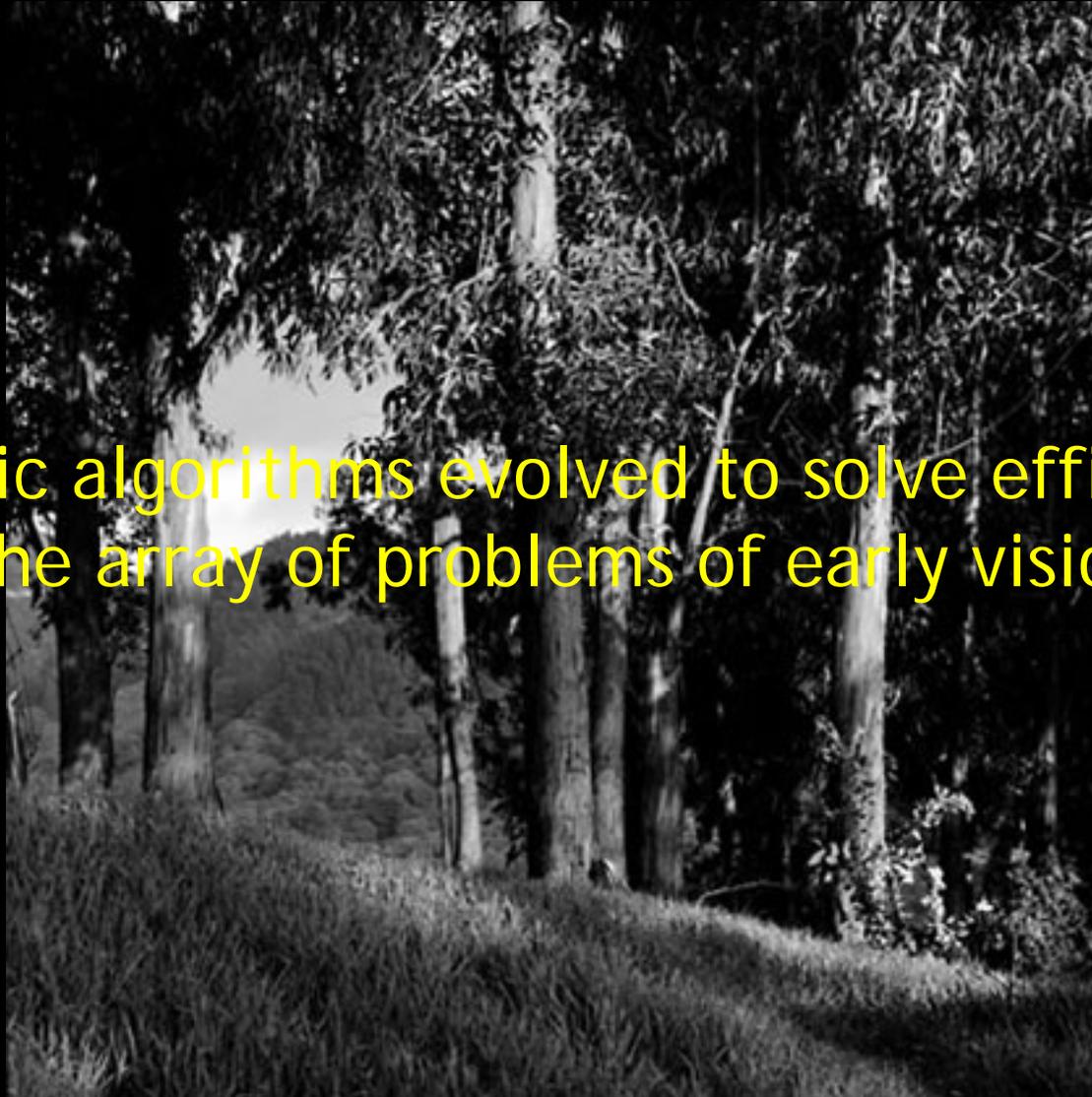


I love you

Let's do lunch

Natural visual statistics

Specific algorithms evolved to solve efficiently
the array of problems of early vision

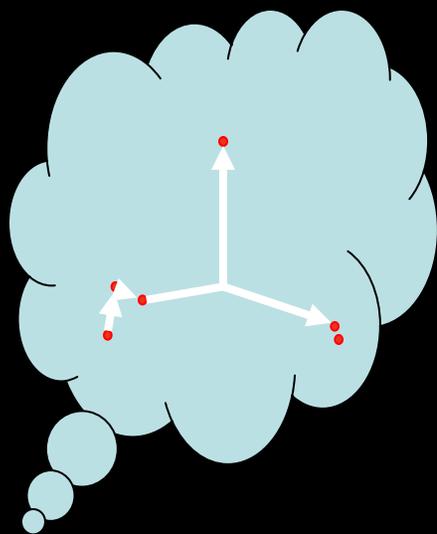


The real world imposes another
difficult requirement on all mobile organisms

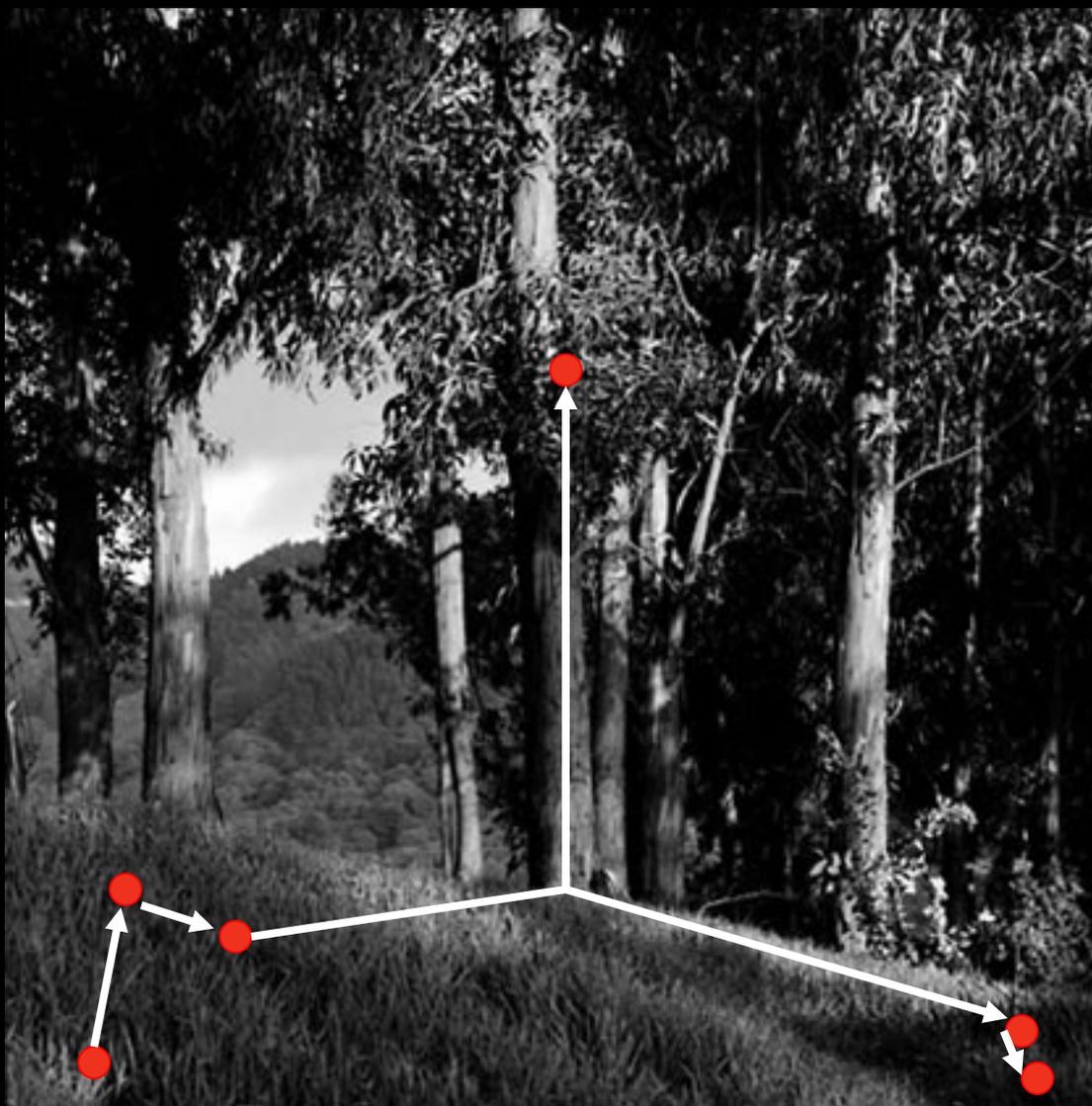


Recharge or die

Natural reward-harvesting statistics?



generic
woodland
creature



What computations should we expect?

Reward harvesting is an economic problem
involving both valuation and choice

Models of reinforcement learning
can provide insight

Goal-directed choice

Select goal

Sustain goal

Pursue goal

guidance
signals

Actual experience

Counterfactual experience
'what could have been'



Midbrain dopamine neurons

Pause, burst, and 'no change' responses represent reward prediction errors - ongoing emission of information



burst
↑
—
↓
pause



naive



After learning



After learning
(catch trial)

$$\text{ERROR SIGNAL} = \text{current reward} + \gamma \text{ next prediction} - \text{current prediction}$$

Can we detect a reward prediction error signal in a human subject?

Neural correlates of these error signals have been observed in conditioning and decision tasks

Passive conditioning tasks

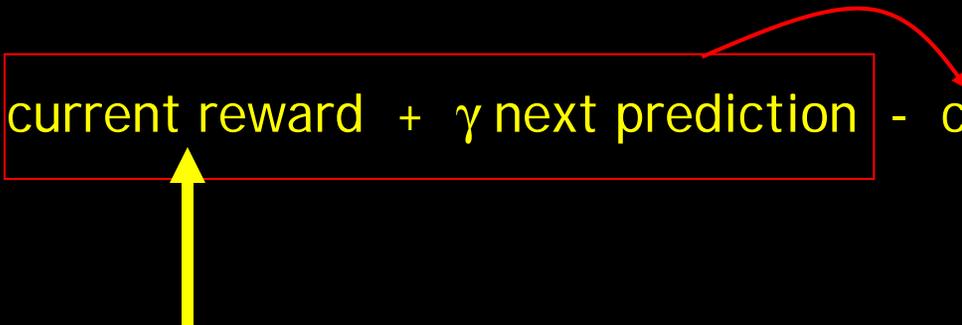
Instrumental conditioning tasks

Sequential decision tasks

Social and economic exchange tasks

Midbrain dopamine neurons

burst and pause responses encode reward prediction errors

$$\text{ERROR SIGNAL} = \text{current reward} + \gamma \text{ next prediction} - \text{current prediction}$$


"I want to solve Fermat's last theorem"

Can these systems be re-deployed for abstractly defined rewards (ideas)?

What about the something more
abstract like the expression and
repayment of trust?

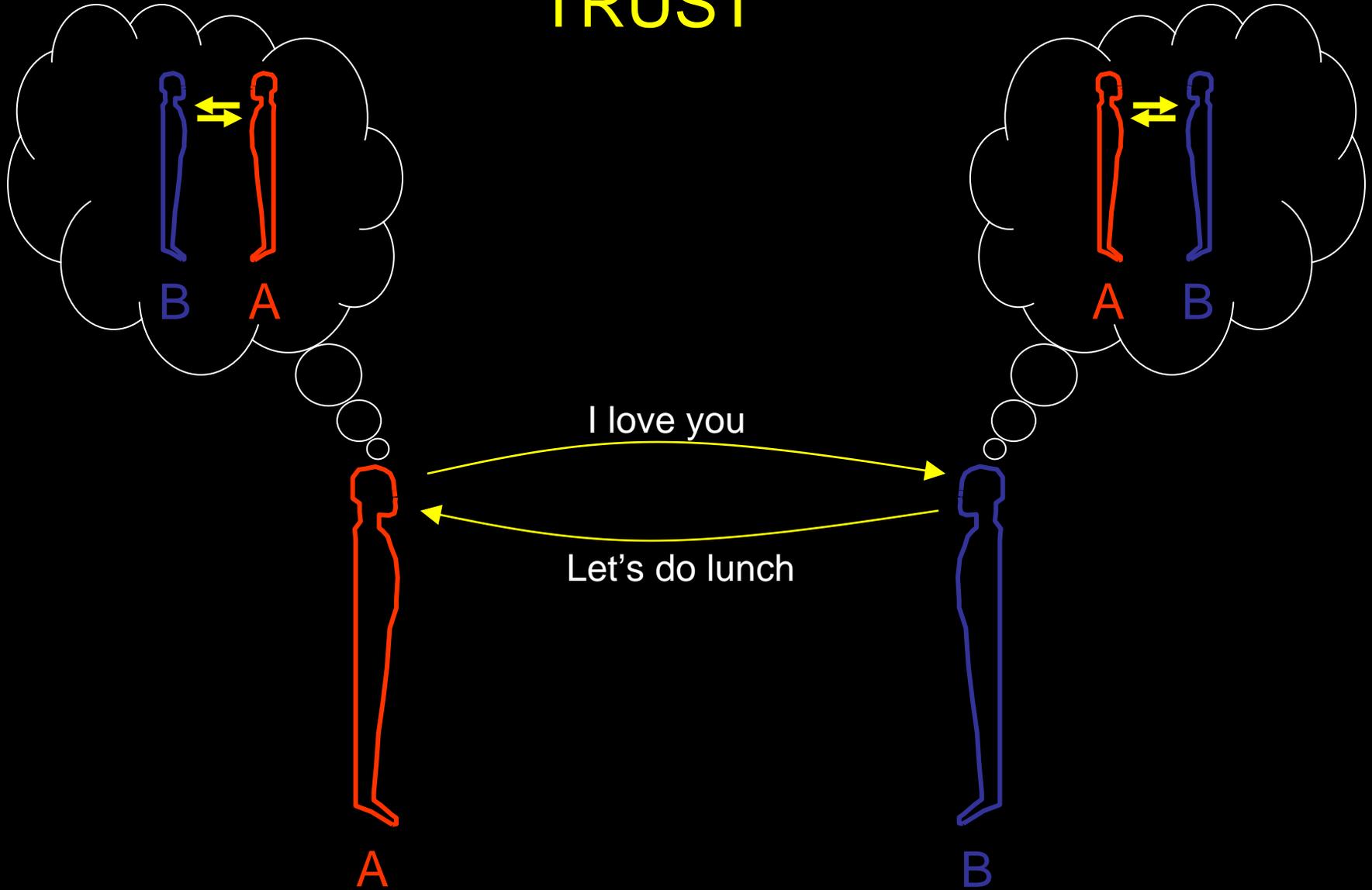
Trust

Modeling

Must involve risk (uncertainty)

Harvesting returns from another agent

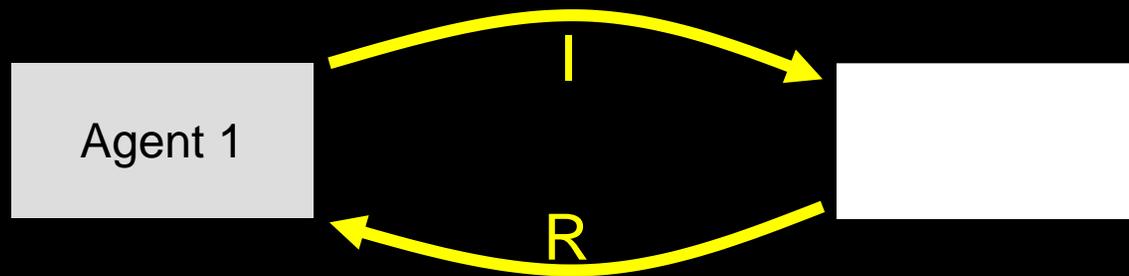
TRUST



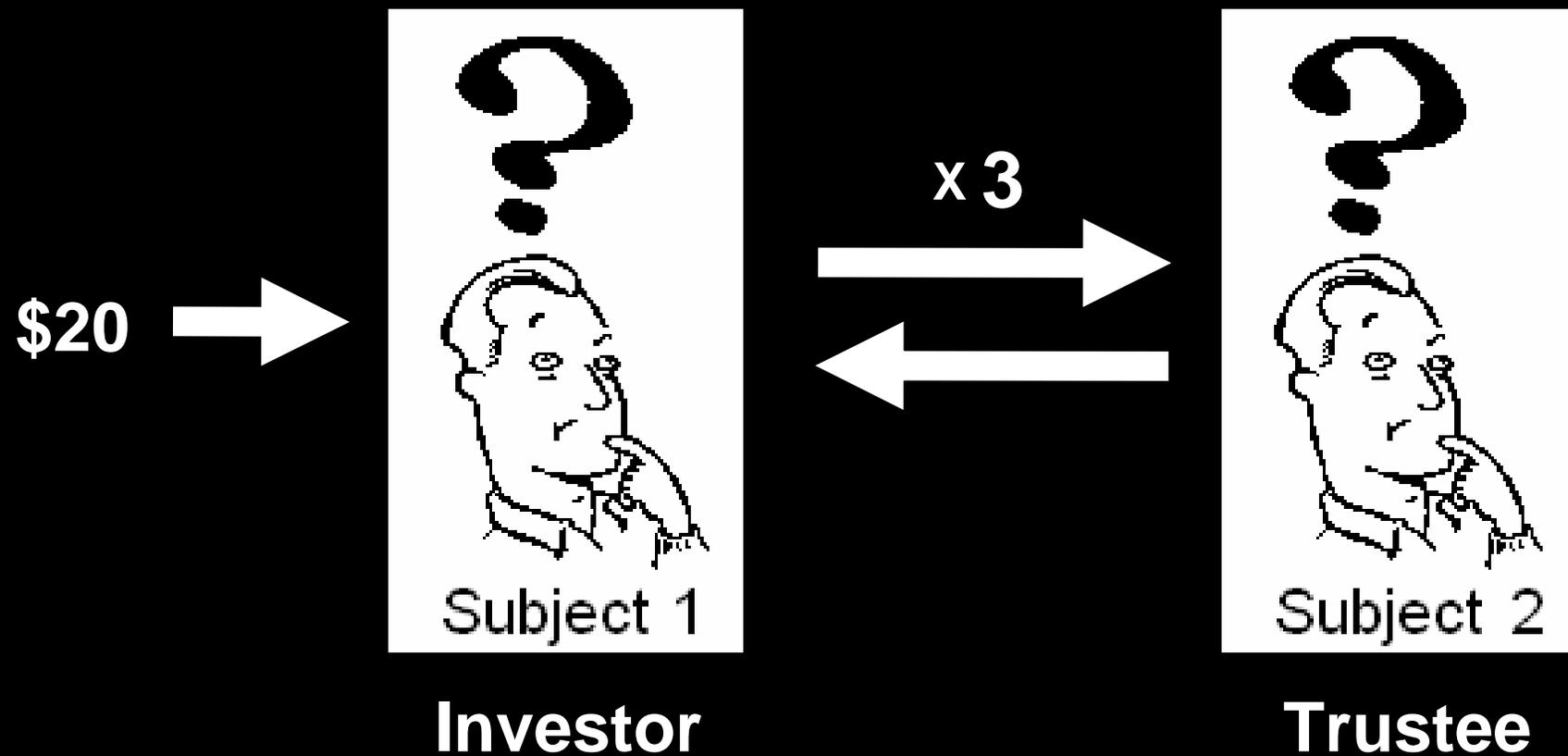
Simplifying and quantifying Trust

(Berg et al., 1995; Weigelt and Camerer, 1988)

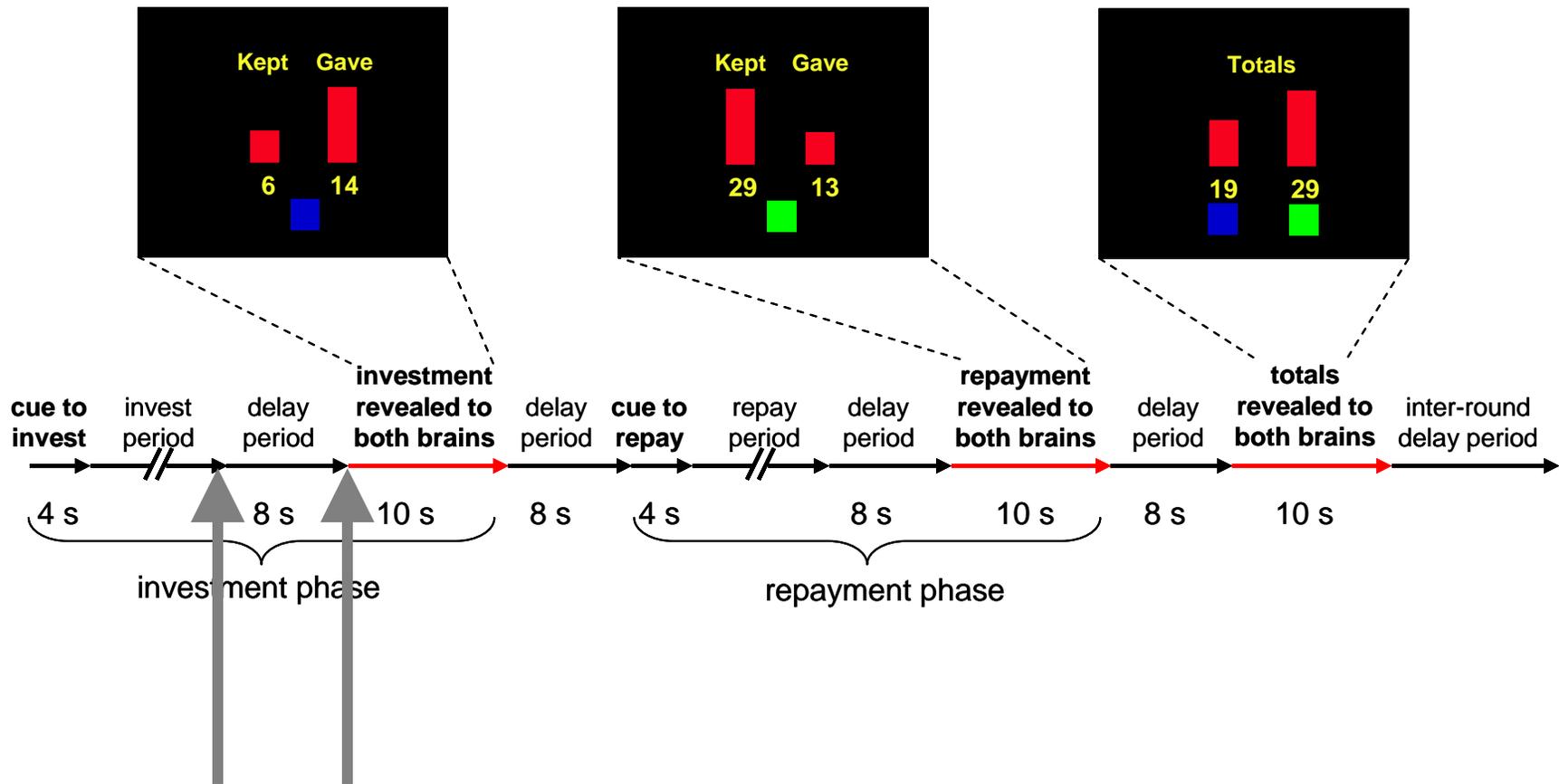
Trust is the amount of money a sender sends to a receiver without external enforcement.



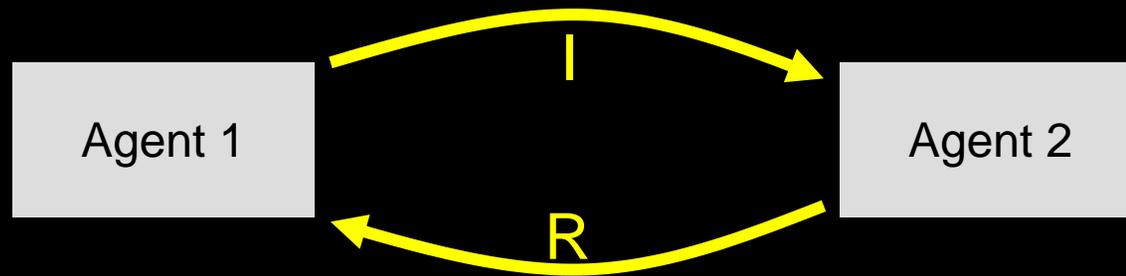
A dynamic version of the Trust game (10 rounds)



Structure of a round

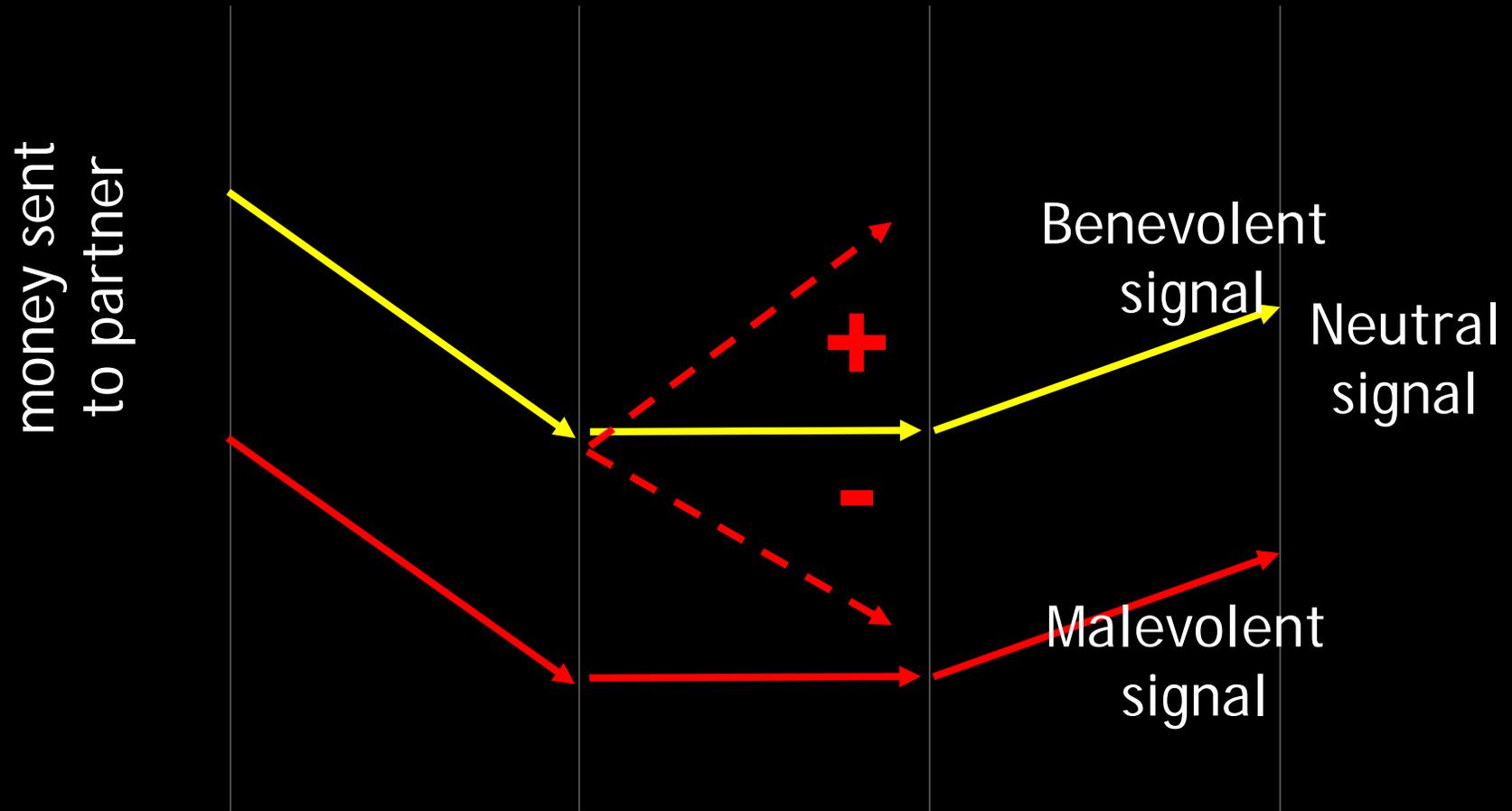


What is the behavioral signal that most strongly influences changes in trust (money sent) ?



Reciprocity = TIT-FOR-TAT

Reciprocity = TIT-FOR-TAT



Questions about brain response

Surprise response

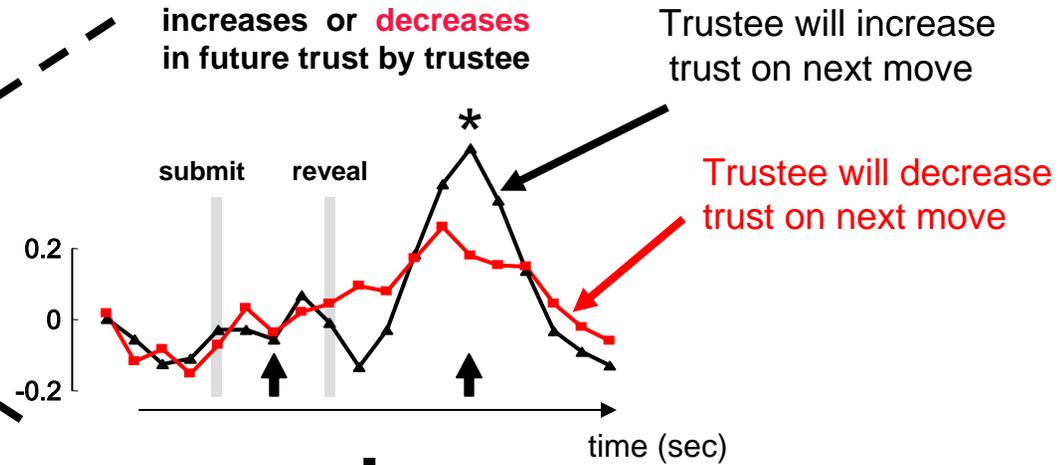
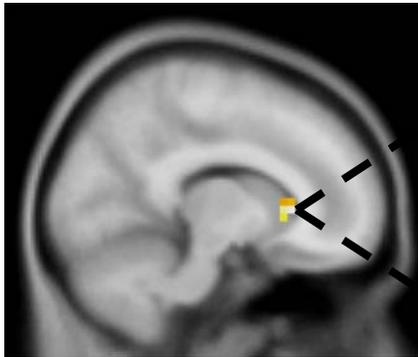
Responses that differentiate benevolent from neutral?

Responses that differentiate malevolent from neutral?

Responses that differentiate benevolent from malevolent?

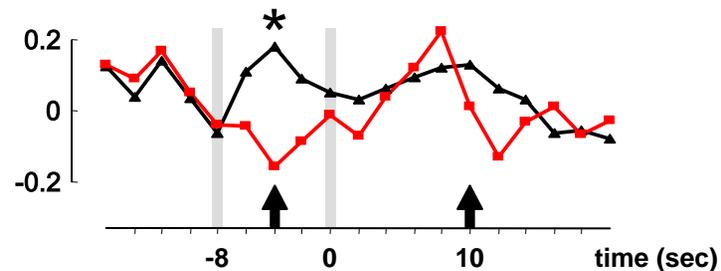
Trustee Brain:

'intention to increase trust' shifts with reputation building

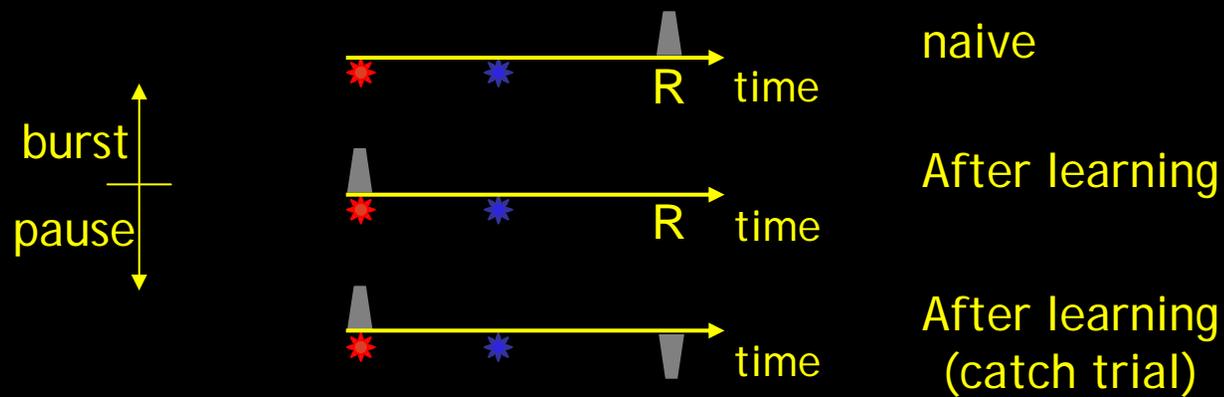


Reputation develops

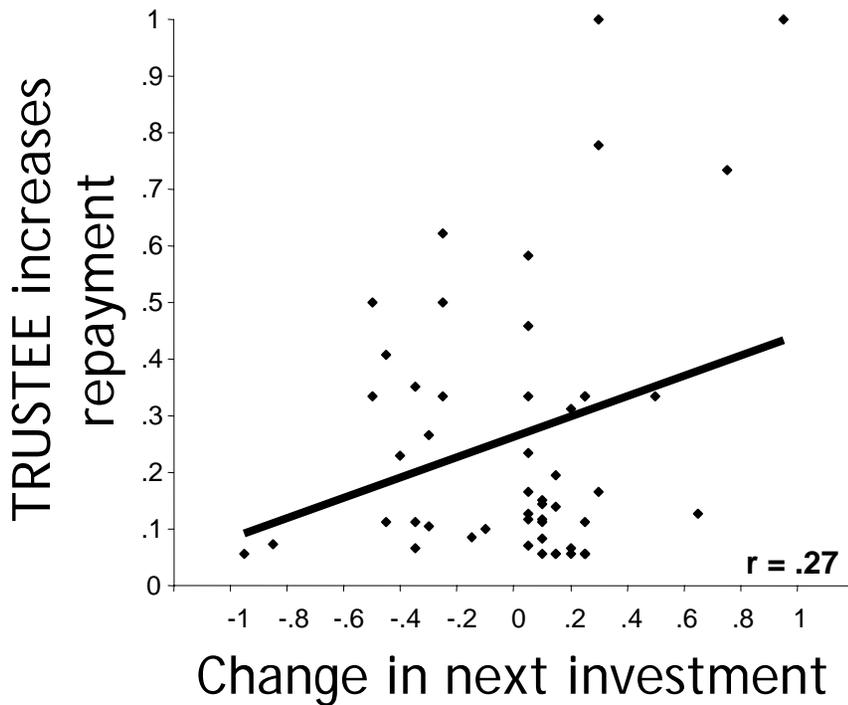
Signal is now anticipating the outcome



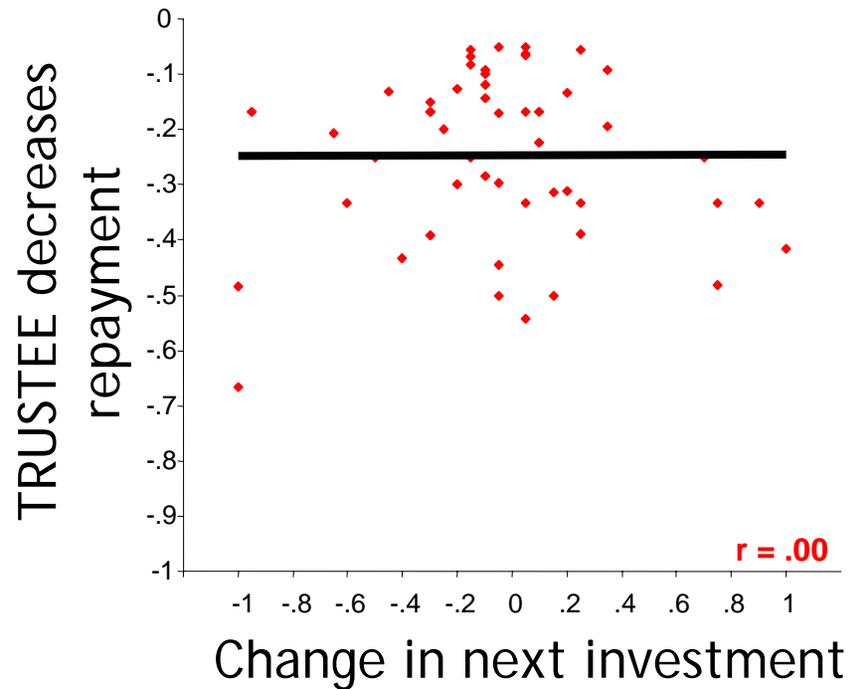
Temporal shift resembles value transfer in reward learning experiments



Why should intentions to increase repayment burn more energy?



Trustee decreases →
positive correlation



Trustee decreases →
No information

Trust?

1. Social trust is about modeling others - always some underlying currency.
2. Re-deploy reward-harvesting machinery that we share with all vertebrates (abstractions gain reward status)
3. Use to probe pathologies (addicted state, autism spectrum disorder, borderline...)

Collaborators

Baylor College of Medicine

Pearl Chiu

Amin Kayali

Brooks King-Casas

Terry Lohrenz

Sam McClure (Princeton)

Read Montague

Damon Tomlin

Caltech

Cedric Anen

Colin Camerer

Steve Quartz

UCL

Peter Dayan

Nathaniel Daw

Salk Institute

Terry Sejnowski

Princeton University

Jon Cohen

Emory University

Greg Berns

University of Alabama

Laura Klinger

Mark Klinger

Families of autistic subjects

Funding Sources

NIDA

NIMH

Dana Foundation

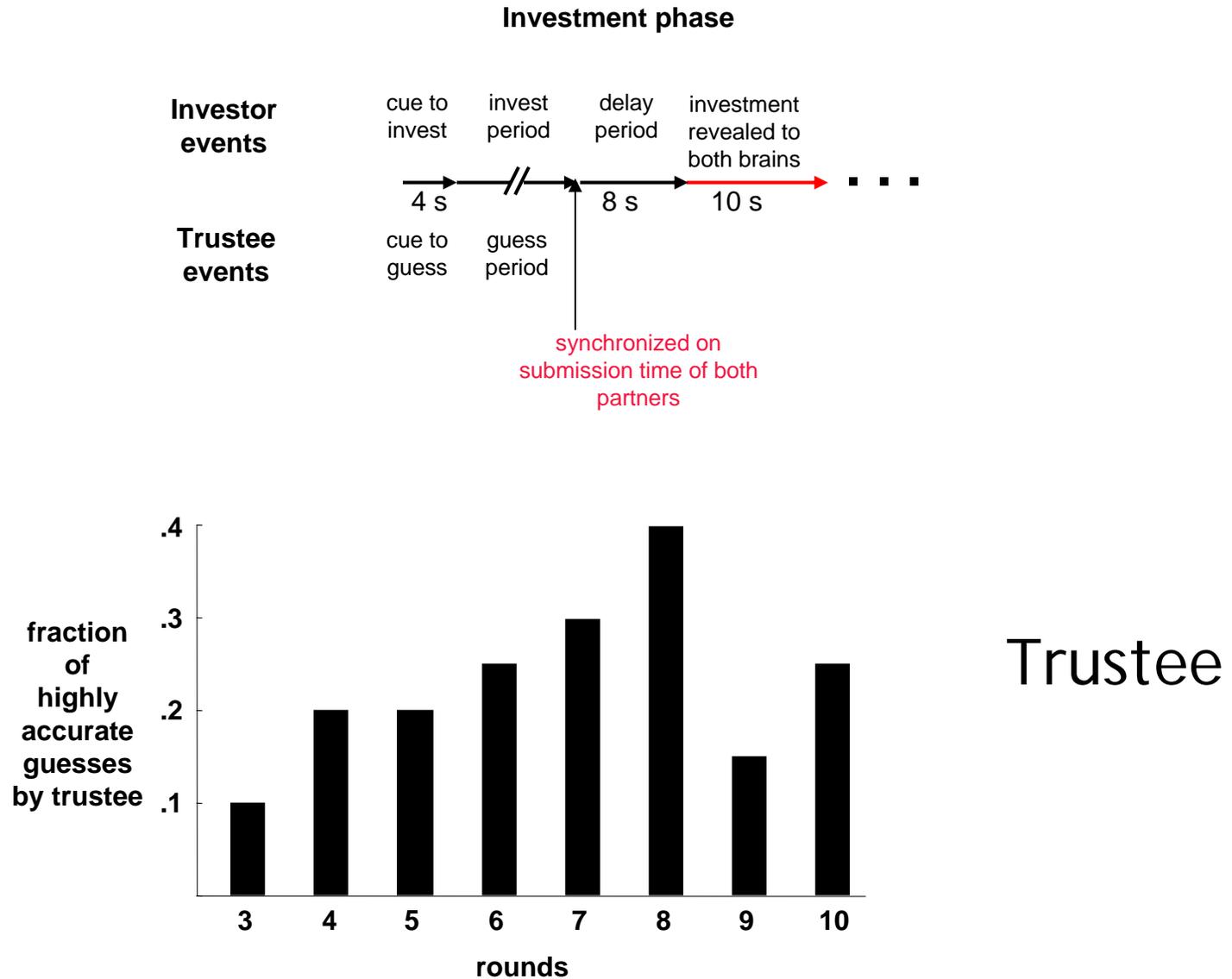
Kane Family Foundation

Angel Williamson Imaging Center

Institute for Advanced Study, Princeton NJ

www.hnl.bcm.tmc.edu/trust

How do we know a reputation is forming?



What is Fair?

Investor



Trustee



Split 'loaf'



Split the profit



Split the total
10 + i each



collective ownership? common goods?